

تقييم صدق وثبات اختبار القبول وقياس إدراك الطلاب:

نموذج جامعة الدلتا للعلوم

د/عبد الرحمن السيد العدل

مناهج وطرق تدريس اللغة الانجليزية

جامعة الدلتا للعلوم والتكنولوجيا

تهدف الدراسة الحالية إلي كشف مدي صدق وثبات اختبار اللغة الإنجليزية للقبول واستكمال إجراءات التقدم لجامعة الدلتا للعلوم والتكنولوجيا, وقياس مدي إدراك الطلاب واتجاهاتهم تجاه هذا الاختبار. واتبعت الدراسة في هيكلها التجريبي علي تتبع خطوات وإجراءات الاختبار المطبق علي الطلاب من حيث: تحديد أهداف الاختبار, محتويات الاختبار, أساليب المراجعة والتقييم, تحديد نظام التقييم, إجراء التحقق من صدق الاختبار (صدق المحتوى والبناء) وثباته في القياس عن طريق استخدام (splithalf and Cronbach's alpha).

وقياس ادراك الطلاب واتجاهاتهم عن طريق تطبيق استبيان قياس الرأي وإجراء مقابلة مع بعض محاضرين اللغة الإنجليزية في الجامعة.

توصلت الدراسة , بناء علي نتائج تحليل صدق الاختبار وثباته وأيضا نتيجة الاستبيان والمقابلة, إلي ضرورة إجراء بعض التعديلات والإضافات علي بعض محتويات الاختبار لضرورة التوافق مع الفروق الفردية في مستويات الطلاب والتلاؤم مع احتياجاتهم تمهيدا لانخراطهم في مناهج ومحتوي اللغة الإنجليزية لمتوقع تدريسه لهم في كليات الجامعة.

الكلمات الدالة: صدق الاختبار, ثبات الاختبار, اختبار القبول, مستوى اللغة الإنجليزية, الاتجاه

Investigating Validity, Reliability and Perception of English Language Placement Test: The Model of Delta University

Abdelrahman E. AlAdl

Assistant Professor of TEFL

Delta University for Science and Technology
Egypt- Gamasa, International Road

Abstract

This study traced the steps that were followed by Delta University for Science and Technology (DUST) to construct an English proficiency test with a high reliability, appropriate validity and strong correlation. The study outlined the procedures used for establishing the language test: determining the aims of the test, constructing test items, estimating and adjusting test items, and carrying out validity (content and face validity) and reliability (split half and Cronbach's alpha). At last, exploring students' and teachers' perceptions of the current English placement test at DUST by administrating a questionnaire and conducting an interview. The assigned questionnaire for students was based on the qualities of a good language test, namely, interactiveness, practicality, reliability, validity and impact. The results indicated that the current placement of Delta University was valid and moderately reliable in benchmarking freshmen students. Regarding students' perception, the result indicated that students perceived the way of comparing with each other in their ability to perform a task, which seemed to create a positive attitude toward language learning. Finally, the results of teacher interviews presented some recommendation for improving the current placement test.

Key words: *test validity, test reliability, placement test, language proficiency, perception*

Abdelrahman E. AlAdl

Investigating Validity, Reliability and Perception of English Language Placement Test: The Model of Delta University

by

Abdelrahman E. AlAdl

Assistant Professor of TEFL

Delta University for Science and Technology

Egypt- Gamasa, International Road

Introduction

Language placement test is a crucial procedure and become a primary requirement of every educational institution. The aim is to benchmark and classify students into their suitable language proficiency level, so that they do not spend time learning materials below or above their levels. It also provides teachers with significant indications that guide them to prepare teaching materials for students of different levels (Illinois, 2012). Delta University for Science and Technology, as an academic institution, has created its own model of English placement test. The high costs of tests, the limited time for students' enrollment, and the variation of the needs of some language programs are reasons for the university not to apply any of specified tests which are available in different institutions, such as IELTS, TOEFL, APTIS...etc.

This study attempted to examine the arrangements that were followed by the university to construct a language proficiency test with a relevant validity, high reliability, and strong correlations to established version of standardized tests. First, this study illustrated the steps that were undertaken to design the four test components (vocabulary, grammar, reading, and listening). Next, the procedures that were used to explore validity and estimate reliability. Then, evaluating the students' and teachers' perception of the test, by conducting relevant questionnaire and an interview.

Finally, the current study discussed the recommended modification on the current test for a better appraisal of the language ability of Delta university students.

The basic question regarding the placement tests is related to the reliability and validity of those tests. Students are enrolled in a grade based on their scores in the placement test. This increases an apprehension regarding its validity, does it assess what it is supposed to test (Brown, 2004). In other words, does it measure students' actual level of English language proficiency?

The placement of English proficiency planned for use in university entrance, so it is important for the test to be distinguished among a suitable range of student proficiency levels. At the same time, the program is intended to make a positive role to English-language learning and teaching in the university by providing useful feedback to test takers beyond the usual pass/fail judgments associated with the university admission test. It is also important to have the test valid and reliable to give the true image of students' proficiency levels, moreover, to investigate the general attitude of both teachers and students towards it to get the proper feedback which may lead to some modifications.

1. Literature Review

Performing placement test to distribute students into relevant foundation courses has been widely covered in the literature. Many studies that have highlighted the basic role of placement test by placing students in the suitable levels that match their needs (Maxwell, 1997& McCabe, 2000). In a study by Al-Adawi & Al-Balushi (2016), they investigated the extent to which the English Language placement test used at Colleges of Applied Sciences (CAS) is achieving this aim by exploring teachers' and students' perceptions of this test via questionnaires and interviews. Furthermore, they

tried to explore the format and content of the test and students' score against their mid-term scores. It was found that face validity of CAS PT ranged from low (teachers' perception) to moderate (students' perception). The majority of teachers and students asserted the importance of including the listening, speaking and reading components in the test. In addition, the study of Sims, (2015) presented the steps a university in Taiwan followed to design a language placement test with a high reliability, appropriate validity and strong correlation to the (TOEFL) test. The study outlined the six steps that has employed for improving the English language test: (1) determining the aims, (2) designing test specifications, (3) building test items, (4) assessing and estimating test items, (5) identifying scoring system, and (6) carrying out validity (content, construct, and concurrent) and reliability (splithalf and Cronbach's alpha).

The research paper of Belfield& Crosta, (2012) made use of student-level data from a statewide community college system to investigate the validity of placement tests and its role for providing information in predicting course grades and raising college performance. They showed that placement tests didn't yield strong predictions of how students will perform in college. So they analyzed correctness rates and four validity metrics for placement tests. They found high "severe" error rates using the placement test cutoffs. The rigorous error rate for English is 27 to 33 percent; i.e., three out of every ten students is severely misassigned. They also showed that using high school grade point average instead of placement tests reduces the severe error rates by half across both English and math.

The study of Armstrong, (2000) tested whether the placement tests of California law were a valid method to design course grades in English and Mathematics. The study constructed a model that explained the variance in test scores

and course grades by checking test scores, student background, and instructor grading variance. The results showed that student characteristics and demographic factors indicated their success. Academic standards are imposed on students by the institution and are not determined by the students' abilities.

Fulcher (1997) investigated face validity and criteria for scoring language placement tests. The results of face validity indicated that the majority of the 71 students who answered the survey expressed that the test was fair enough, however, some students expressed their anxiety regarding the ambiguity of some test items. Four students requested longer tests while few others wanted subject-specific content. Moreover, there was a general worry about the exam room environment which does not support reliability of results. In addition, Shin and Kin (2006) investigated the reliability and validity of the design, evaluation and results analysis of a language placement test. The findings showed significant results in terms of reliability and validity of the test which consist of two reading tasks (worth 12 marks and should be completed in 20 minutes) and a descriptive writing task (20 minutes). Generalization of the study results could not be accomplished due to the limitation of the study such as small number of participants and the likability of interference of practicality issues.

2. Statement of the Problem

The current study basically attempted to examine content and face validity of the English placement test that have been used at DU. The aim is to suggest the relevant improvement for the test regarding teachers' and students' perception. There is a need for improving the annual placement test that assigned for freshmen students (admission test). The aim is to accurately assess the

proficiency of students, furthermore to shed the light on planning for English courses.

3. Delimitations

The study was performed at DUST, in the academic year 2017-2018, tackling the assigned version of English language placement test. The participants of the study were sixty students and eight English lecturers in the university.

4. Research Questions

The study was conducted to answer the following questions:

- (1) To what extent is the English language placement test used in DUST, valid and reliable in assessing students' true language proficiency level?
- (2) What are teachers' and students' perceptions of the current English placement test?

5. Methodology

The study follows the procedures for developing the language exam: determining the aim of the test, constructing test items, evaluating and revising test items, specifying scoring procedures, and performing validity (content, concurrent and face) and reliability (split half and Cronbach's alpha). Finally, exploring teachers' and students' perceptions of the current English placement test.

A. Participants and Instruments

This study was carried out at (DUST) for Science and Technology among the enrolled freshmen students in the academic year 2017- 2018. Sixty students participated in the study by answering the questionnaire (Appendix B) to determine students' perception of the placement test. In addition, eight of the university English lecturers were interviewed (Appendix C) to state their attitude towards the placement test layout and content and whether they found it

valid enough to place students in the right proficiency language level.

B. Procedures and data collection

Exam Construction

1- Determining the purpose: The first purpose was to place students into different levels of Freshman English for classes based on their language ability. Secondly, to create a diagnostic tool to identify students' weaknesses and strengths. The test is composed of 60 multiple-choice questions, and scores were calculated electronically. Each question had four choices, but only one correct answer. The time allotted for the test was 60 minutes.(Appendix A)

2- Designing test specifications: the test is composed of four constructs: Grammar, Vocabulary, Reading, and Listening. The Grammar Section (25%) is composed of 15 questions for a total of 15 points. The Reading Section (25%) is composed of three short passages with 5 questions per passage for a total of 15 points. The Listening Section (25%) is composed of two parts: Short Dialogues (8questions), and Appropriate Response (7 questions). the Vocabulary Section Vocabulary-in-Context Questions: Fifteen questions asked students to find the synonym that made the most sense when it was substituted for the word or phrase in question.

3- Evaluating and revising test items: Item difficulty was used to ensure a near normal distribution of scores, to make each subsection of the exam progressively more difficult, and to create an overall total mean score of between 55-60%. This meant that some items were easy while others were difficult. But, the total mean of all items would range between 55-60%. For a 60 point exam, this would ensure near normal distribution of scores. Item discrimination was

used to differentiate appropriately between high and low test takers. Distractor analysis was used to make sure that distractors were efficiently distributed. Each distractor was plausible, but incorrect, and had some respondents. For ideal items, 55-60% of the students selected the correct answer (item difficulty of 0.55 to 0.60) and each distractor was selected by 10-25% of the test takers. For all items, students in the lower group (bottom 27% of all test takers) than in the upper group (top 27% of all test takers) selected distractors. In other words, every distractor was chosen by a higher percentage of low achievers than high achievers. This was done to ensure that the higher level students were not being drawn to an incorrect answer by poor distractors.

4- Specifying scoring procedures

The Grammar Section (25%) was composed of 15 questions for a total of 15 points. The Reading Section (25%) was composed of 15 multiple choice questions for a total of 15 points. The Listening Section (25%) was composed of two parts: Short Dialogues (8 questions), and Appropriate Response (7 questions) for a total of 15 points. The test is computer based sheet, which automatically corrected and scores is calculated by using a soft system program.

5- Administering students' questionnaire and teachers' interview:

In order to explore students' and teachers' perception of the placement, a questionnaire was conducted after the test. Quantitative values are to be assigned to the Likert scale to analyze the data. The values were calculated including mean scores, standard deviation and percentage agreement for all the items. Comparisons of mean scores and standard deviation were analyzed. In addition, a semi-structured interview with teachers in the university was conducted to investigate their perception of the placement test (see Appendix, b&c).

6. Data Analysis and Results

Exploring Test Validity

The validity of a test, or the extent to which a test actually measures what it is initially intended to measure, is a complex criterion in the field of testing (Brown, 2001).

In order to determine whether the placement test was an appropriate instrument, three ways were used to explore the validity of the test. First, a content validity study was carried out to investigate all test items. Second, a cross-comparison correlation study between the assigned test and other established standardized exams was performed to examine its concurrent validity. Third, Face Validity was carried out through interviewing eight English language instructors to explore their perceptions of the current placement test and suggestions for improving it.

1- Content Validity: A content validity study was conducted based on a comparison of test specifications and test content. Following Hughes' (2003) recommendations, these comparisons were made by eight English instructors who were trained in instruction and testing, but were not engaged in the construction of the test. These instructors reported that the test items were appropriate measures of the desired test specifications for grammar, reading, vocabulary and listening.

2- Concurrent Validity: The results of a cross-comparison correlation analysis of 50 freshmen who took both the current placement test and the IELTS showed that the total scores of the current test had a high correlation ($r = 0.87$) with the total scores of the IELTS. High correlations were reached: the listening section ($r = 0.83$), the reading section ($r = 0.82$) and grammar/vocabulary sections ($r = 0.83$).

A similar analysis of 500 freshmen who took both the current placement test and the Secondary Stage Exit Exam (Thanaweya Amma) showed that the total scores of the placement also had a significant correlation ($r = 0.77$) with the total scores of the Secondary Stage Final Test. The correlation indicated that the placement has strong concurrent validity similar to that of the IELTS and moderate concurrent validity to the Secondary Stage Exam.

3- Face Validity: Face validity concerns the form of the test. The face validity of the current placement test was measured by interviewing eight English instructors and by administering a questionnaire to the students.

Exploring Test reliability

Reliability is the degree of consistence of the test in measuring whatever it does measure (Sims, 2015). Reliability analysis was conducted to assess the test-retest reliability (Durrant, 2014). A split-half method was used to evaluate the content reliability of the current placement test, while a Cronbach's alpha approach was used investigate the item variance reliability. For the split-half method, the exam was divided into two equivalent halves with each half composed of matching content, or skills. For example, each test item was carefully matched with a similar type of question from the other half. Questions that dealt with the main idea of paragraphs were paired up with other questions designed to measure the understanding of the main ideas of paragraphs. The same was done for comprehension/details questions, vocabulary in context questions, and inference questions. As for the grammar section, similar grammatical points were paired together. A similar procedure was followed for the listening section. The Spearman-Brown split-half reliability coefficient was calculated to be $r = 0.869$, while Cronbach's alpha reliability coefficient was $r =$

**Investigating Validity, Reliability and Perception of English Language
Placement Test: The Model of Delta University**

0.878. In short, the placement test could be considered a reliable instrument based these two high reliability coefficient.

- Students and Teachers' Perception of the placement test

a) Students' Questionnaire: (a modified version of Kohonen's Authentic assessment in affective foreign language education (Kohonen, 1999). The questionnaire items have been grouped into language test qualities (Bachman & Palmer, 1996)

(Appendix B)

No	Item	Mean	Agree %	ST.D
1	When I take a test in English, I focus on only one correct answer.	3.22	81	.591
2	Time limits do not allow me time to finish my test.	3.06	77	.593
3	Multiple choice exams have similar items to those I studied in class.	2.93	73	.575
4	I have a positive attitude towards language learning.	2.95	74	.500
5	Tests determine what students cannot do.	2.37	63	.506
6	We should be tested on what we know to boost motivation	2.97	78	.508
7	Multiple choice exams test only lower order knowledge.	3.04	79	.717
8	The test evaluated my ability to perform a task.	2.96	76	.506
9	The students should do a practice test before the actual test.	3.31	88	.768
10	The test offers the student a variety of different items.	2.54	68	.810
11	The current English test is stressful.	3.02	76	.487

Abdelrahman E. AlAdl

12	The test was a relative competition (you win, I lose).	3.03	77	.519
13	By taking the test, I know what I can and cannot do in English	2.92	7 3	.686
14	Test items simulate real-world tasks.	2.95	7 4	.531
15	The test teaches students why they fail.	2.76	71	.548
16	It is embarrassing to speak in English, in speaking tests.	3.03	7 6	.510
17	Students were compared with each other.	2.91	74	.532
18	I worry about making mistakes.	2.97	73	.689
19	The test allowed me to think and respond to a question.	2.94	7 4	362
20	Active awareness of learning was promoted.	2.99	7 5	.643
21	There was more than one correct answer	3.08	77	.731
22	My test score reflected my abilities in English	3.08	7 7	.630
23	The current English test forbids students to interact.	3.11	78	.810
24	The test determines my strengths and progress in English	3.03	76	486

The results of the above table showed some points that were related to students' background learning. This learning characteristic was reflected in items 11, 16, and 18. the students were embarrassed to speak in English (item 16, 76% agreed). In item 18 (73% agreed), they worry about making mistakes, which is characteristic of passive learning in the Egyptian classroom as students are generally seeking an answer and not a conversation. In item 11, (76% agreed), they view current English tests as stressful. This stress resulted from the competitive environment that the students

have to deal with on an ongoing basis if they are to be successful (selected) for university admission.

Item 24 “emphasized their strengths and progress in English” had 76% agreement. This showed that the students have a positive attitude towards language learning. Item 13 “determine what he can and cannot do in English” received 73% agreement and this proposed that students can measure their abilities. That the test task “allowed me to think and respond to a question” (item 19) received 74% agreement. These results do suggest, however, that motivation is related to cognitive constructs, self-efficacy in a language task, and language learner's constructions of self trust (Bachmann, 1990). By evaluating a student in a qualitative manner, a teacher is able to praise the strengths as well as constructively criticize weaknesses.

To keep reliability on a language test, a large number of items usually ensures a greater range of scores that effectively separates candidates (Qian, 2007). Item 10, “the test offers a variety of different items,” achieved (68%) agreement. The reason may be that students realize that the criteria for correctness is limited in a multiple choice exam as usually only one correct answer is possible. If there is no penalty for guessing, then they can simply engage in an elimination process among a few items to arrive at the correct answer.

Regarding the test validity, which is the agreement between attempts to assess the same attribute through two different methods (Bachman & Palmer 1996). The positive response of students regards item 22 (77% agreement) “the tasks were appropriate given my ability in English.” implied that both the content and the way of testing were valid considering the aims of the course.

To insure content validity, a practice test is warranted. The students mildly agree (88%) that “students should do a practice test before the actual test” (item 9). They may show that some students can develop a better ability to answer unfamiliar questions that are not content embedded. This may be due to the ability to achieve better depending on the number of times the test is taken.

Summary of students' perception

- The students stated that they were active as opposed to passive learners (item 20). This may be because after finishing the placement test, they could predict the required language skills.
- They reported that they had positive approach towards language learning (item 4) and that they should be tested on what they know to boost motivation (item 6).
- They admitted that they were being assessed on their ability to complete tasks (item 8).
- The students asserted that the current test was being a reliable indicator of their ability (item 13).
- Score interpretation seems more relevant to the students as the test was an interactive activity using real world situations. (Item 14).
- They perceived the test tasks as being appropriate for their abilities (item10). This implies that they are able to determine their ability and content mastery...

b) Teachers' Perceptions of the placement test (Face Validity)

Six English language foundation teachers were interviewed to determine their perception of the current placement test and recommendations to improve the test. They are experienced in teaching English and assessment. Some of them suggested including speaking components in the test as they are important skills. Another teacher

recommended modifying the reading component to make it a better judgment of students' reading ability rather than a measure of grammar or vocabulary. Another one stated that the writing section should be added to the test uses familiar topics to the students and the reading text is not long enough to test reading skills and the only sub-skill tested is scanning for details. Moreover, all the teachers encouraged adopting the computerized placement test because it was suitable with the large numbers of candidates. (Appendix C)

7. Discussion and Conclusion

This study attempted to evaluate the placement test used by (DUST) focusing on teachers' and students' perceptions, and the extent of validity and reliability of the current test. The explanation of test content showed that the separation of reading from grammar was found. The separation of listening from reading and grammar is widely accepted (Bae & Bachman, 1998; Shin, 2005; Song, 2008). The reading items focused on main ideas, specific details, and vocabulary in context, and did not require a deep understanding of the grammar. Moreover, the grammar items mostly dealt with appropriate verb tense, subject-verb agreement, and count versus non-count nouns, possessive pronouns, conjunctions, and passive voice.

Regarding the test reliability, both a split-half method and an item variances approach were used to measure the internal reliability of the test. First, as it was designed to measure different language skills (i.e., grammar, reading, vocabulary, and listening), it was reasonable to estimate the internal consistency with a Spearman-Brown split-half reliability coefficient (Bachman, 2004). Second, for the scores of items to be independent and parallel measures with similar variances, a Cronbach's alpha method was also appropriate. Quite simply, the split-half method estimated

the reliability based on the content of the exam, while the Cronbach's alpha reliability coefficient estimated the reliability based on the variance of the individual items. As reported in the results section, both the reliability based on both the "content" ($r = 0.878$) and the "variance" ($r = 0.869$) of the test were high.

The results show that placement test might be useful for students in their later learning after admission. There are many researches that have emphasized the importance of placement test in helping students succeed in their studies later by placing them in the suitable courses that match their needs (Casazza & Silverman, 1996; Maxwell, 1997; McCabe, 2000).

Regarding the students' and teachers' perception, the majority of teachers and students emphasized the importance of including the writing and speaking components in the test. Moreover, a modified version of the reading section of the test needs to be incorporated into the test. It is suggested to design a new test, taking into consideration the findings of this research and pilot it to test its effectiveness.

Questionnaire and interviews point to a belief in the power of the university admission test to influence teaching and learning, to a recognition that a range of language skills is relevant to university study, and to common acceptance of the suggestion that changes in the content of the examinations should lead to changes in the content of lessons.

The majority of teachers and students agreed that changes to the tests would encourage changes in the focus of teaching and learning, it might be expected that teachers and students would strongly endorse moves to introduce tests of reading, listening, and writing. These points agreed with the finding of (Bachman & Palmer, 1996; O'Sullivan, & Weir, 2011 and Weir, 2005)

The current placement test of Delta University has met the factors of proficient test, that was set by Brown (2004); a practical exam: (1) is not too expensive, (2) remains within appropriate time constraints, (3) is relatively easy to administer, and (4) has a scoring/evaluation procedure that is specific and time-efficient (p. 19).

8. Recommendation and Implications

Overall the majority of students and teachers believed that the placement was reliable and valid. However in order to provide more information regarding the appropriateness of student placement it is recommended that additional quantitative research be conducted. Specifically, it is recommended that correlations between placement test scores and success in courses be examined to determine the degree to which placement scores are correlated with student success in courses. The few limits of this study may be concluded as follows:

- 1- Students may score low in the test due to practicality issues, physical factors during the exam such as stress, heat and sickness, which this paper has not investigated.
- 2- Comparing students' previous scores against the placement test scores might not be the best way of assessing its effectiveness because of the differences between the two exams in terms of purpose, content and format. In addition, the period between the two tests.

Reference

1. Al-Adawi S. A & Al-Balushi A. K. (2016). Investigating Content and Face Validity of English Language Placement Test Designed by Colleges of Applied Sciences, *English Language Teaching*; Vol. 9, No. 1; Pp. (107-121).
2. Armstrong, William B. (2000) The Association Among Student in Courses, Placement Test Scores, Student Background Data, and Instructor Grading Practices. *Community College Journal of Research and Practice*. 24: 681-695.
3. Bachman, L. F. (2004). *Statistical analysis for language assessment*. Cambridge, England: Cambridge University.
4. Bachman, L. & Palmer, A.(1996). *Language testing in practice: Designing and developing useful tests* (pp. 17-42). Oxford: Oxford University Press.
5. Bae, J., & Bachman, L. F. (1998). A latent variable approach to listening and reading: testing factorial invariance across two groups of children in the Korean/English two-way immersion program. *Language Testing*, 15, 380-414.
6. Belfield C. R. & Crosta P. M. (2012) Predicting Success in College: The Importance of Placement Tests and High School Transcripts, *CCRC Working Paper* No. 42, Teachers College, Columbia University.
7. Brown, H.D. (2001) *Principles of Language Learning*. 4th ed. White Plains, New York: Longman.
8. Brown, H. D. (2004). *Language assessment principles and classroom practice*. White Plains, NY: Longman.
9. Durrant, J. (2014) "The Reliability and Validity of the Ten Test and Exploring a New Visual Version". M.A. thesis, McMaster University, Hamilton, Ontario.

10. Flucher, G. (1997). An English Language Placement Test: Issues in Reliability and Validity. *Language Testing*, 14(2), 113-138.
<http://dx.doi.org/10.1177/026553229701400201>.
11. Casazza, M., & Silverman, S. (1996). *Learning assistance and developmental education*. San Francisco, CA: Jossey-Bass.
12. Hughes, A. (2002). *Testing for language teachers*. Cambridge, England: Cambridge University Press.
13. Illinois, G. (2012). *English Placement Test (EPT)*. Retrieved on March 2, 2014, from www.publications.uiuc.edu.
14. Kim, H., & Shin, H. (2006). A reading and Writing Placement Test: Design, Evaluation and Analysis. *Teachers college, Columbia University Working Papers in TESOL & Applied Linguistics*, 6(2), 1-36.
15. Kohonen, V. (1999). Authentic assessment in affective foreign language testing. In J. Arnold (Ed.), *Affect in language learning*. (pp.279-294). Cambridge: Cambridge University Press.
16. Maxwell, M. (1997). *Improving student learning skills*. Clearwater, FL: H & H Publishing.
17. McCabe, R. (2000). *No one to waste: A report to public decision-makers and community college leaders*. Washington, DC: Community College Press.
18. O'Sullivan, B., & Weir, C. J. (2011). Test development and validation, *Language testing: Theories and practices*, 13-32. Basingstoke: Palgrave.
19. Qian, D. (2007) Assessing University Students: Searching for an English Language Exit Test. *RELC Journal*; 38; 18. pp. 18-37.

20. Shin, S. K. (2005). Did they take the same test? Examinee language proficiency and the structure of language tests. *Language Testing*, 22, 31- 57.
21. Sims J. M. (2015) A Valid and Reliable English Proficiency Exam: A Model from a University Language Program in Taiwan, *English as a Global Language Education (EaGLE) Journal*: Vol. 1 No.2, Pp. 91-125.
22. Song, M. Y. (2008). Do divisible subskills exist in second language (L2) comprehension? A structural equation modeling approach. *Language Testing*, 25, 435-464.
23. Weir, C.J. (2005) *Language Testing and Validation: An Evidence-Based Approach*. Oxford: Palgrave.

Appendices

Appendix (A)

The placement Test (sample of questions)

	delta university
Started on	Sunday, 6 August 2017, 1:37 PM
State	Finished
Completed on	Sunday, 6 August 2017, 1:56 PM
Time taken	18 mins 47 secs
Grade	48.00 out of 60.00 (80%)

اكتب اسمك الرباعي باللغة العربية

عمر نبيل عبدالستار المرسي

(a) Listening

(b) Question 1

Play Video

What is Noah doing?

Select one:

a. He's swimming

Investigating Validity, Reliability and Perception of English Language
Placement Test: The Model of Delta University

b. He's talking on the phone.

c. He's windsurfing

(c) Question 2

(i) Question text

Play Video

What doesn't the holiday include?

Select one:

a. A hotel near the beach

b. Sightseeing tours

c. Two meals a day at a hotel

(d) Grammar

(e) Question 16

(i) Question text

He ---- very tired, he is falling asleep.

Select one:

a. in

b. am

c. are

d. is

(f) Vocabulary

(g) Question 31

he ---- is very bad today, it is raining cats and dogs.

Select one:

a. weather

- b. whether
- c. air
- d. temp

(i)

(h) Reading

Researches suggest that there are creature's المخلوقات that do not know what light means at the bottom of the sea. They don't have either eyes or ears; they can only feel. There is no day or night for them. There are no winters, no summers, no sun, no moon, and no stars. It is as if a child spent its life in darkness in bed, with nothing to see or hear. How different our own life is! Sight shows us the ground beneath our feet and the heavens above us - the sun, moon, and stars, shooting stars, lightning, and the sunset.

46. Judging from the passage, we can say that this topic is mainly about

- a) life of sea creatures at the bottom of the sea
- b) how changes in the seasons are perceived by the deep-sea creatures
- c) how wonderful our lives were and will be
- d) the differences among creatures of the earth and those of the sea
- e) the superiority of human beings over some creatures in terms of senses

47. We discover that the sea creatures in the story

.....

- a) have the same senses that we do

Investigating Validity, Reliability and Perception of English Language
Placement Test: The Model of Delta University

- b) have no sense of hearing as well as sight
- c) hear the sounds of the ocean
- d) live in darkness because no light reaches to the bottom
- e) do not hear the sound of sea as they are accustomed to it

Appendix (B)
Students' Questionnaire

No	Item	Agree	Disagree
1	When I take a test in English, I focus on only one correct answer.		
2	Time limits do not allow me time to finish my test.		
3	Multiple choice exams have similar items to those I studied in class.		
4	I have a positive attitude towards language learning.		
5	Tests emphasize what students cannot do.		
6	We should be tested on what we know to boost motivation		
7	Multiple choice exams test only lower order knowledge.		
8	The test evaluated my ability to perform a task.		
9	The students should do a practice test before the actual test.		
10	The test offers the student a variety of different items.		
11	Current English tests are stressful.		

Abdelrahman E. AlAdl

12	The test was a relative competition (you win, I lose).		
13	By taking the test, I know what I can and cannot do in English		
14	Test items simulate real-world tasks.		
15	Tests teach students why they fail.		
16	It is embarrassing to speak in English.		
17	Students were compared with each other.		
18	I worry about making mistakes.		
19	The test allowed me to think and respond to a question.		
20	Active awareness of learning was promoted.		
21	There was more than one correct answer		
22	My test score reflected my abilities in English		
23	Current English tests forbid students to interact.		
24	The test emphasized my strengths and progress in English		

Appendix (C) Interview Questions

This survey is part of a research project examining the Effectiveness of English language placement test at the colleges of Applied Sciences. We do appreciate your time and effort in answering this survey. All responses and answers are confidential and information provided will be used for research purposes only.

Name:

Years of experience in teaching:

- 1) Have you looked at English placement test which is run in the college?
- 2) What do you think about it?
- 3) Does it cover all language skills?
- 4) Do you think it represents the students' actual levels of language proficiency? Why?
- 5) Have you ever encountered a student that you thought he/she might be misplaced in that level?
- 6) If yes, how did you reach that conclusion?
- 7) What was your reaction?
- 8) What are the drawbacks of the current placement test?
- 9) Do you think that the placement test should be modified?
- 10) Do you have any suggestions for changing the placement test? What are the suggestions?